

Intrinsically Motivated Multimodal Structure Learning

Jay Ming Wong^{†‡} and Roderic A. Grupen[‡]

[†]Planning, Autonomy, and Automation Group, Draper Laboratory, Cambridge, MA, USA

[‡]Laboratory for Perceptual Robotics, University of Massachusetts Amherst, Amherst, MA, USA

jmwong@draper.com, grupen@cs.umass.edu

Abstract—We present a long-term intrinsically motivated structure learning method for modeling transition dynamics during controlled interactions between a robot and semi-permanent structures in the world. In particular, we discuss how partially-observable state is represented using distributions over a Markovian state and build models of objects that predict how state distributions change in response to interactions with such objects. These structures serve as the basis for a number of possible future tasks defined as Markov Decision Processes (MDPs). The approach is an example of a structure learning technique applied to a multimodal affordance representation that yields a population of forward models for use in planning. We evaluate the approach using experiments on a bimanual mobile manipulator (uBot-6) that show the performance of model acquisition as the number of transition actions increases.

I. INTRODUCTION

Humans accumulate a large repertoire of action-related knowledge from experiences over a lifetime of problem solving. As infants, we explore the world and entities in our environment, building representations for future use through play. We do this because we are inherently curious and discovery is rewarding for its own sake—we are *intrinsically motivated* to acquire models of the world [1–3]. Many researchers have explored intrinsic motivation as a key component for developing curious, exploratory, and autonomous behavior—for instance, in the acquisition of visuomotor skills for robots [4]. Hierarchical approaches have been developed employing intrinsic motivation to learn new skills autonomously [5, 6]. A number of intrinsic motivators have been proposed [4, 7] with approaches in contrast with previous work that relied on hand-built representations tailored to a particular task [8–10]. Our view is that autonomous exploration and intrinsically motivated discovery should prove more robust and transferable than hand built knowledge representations.

Insight from cognitive psychology has influenced many researchers to investigate knowledge representations [7, 11–27]. Among these, the notion of *direct perception* and *affordances* proposed originally by Gibson [28] is particularly relevant to our approach. Gibson’s theory of affordance advocates for modeling the environment directly in terms of the actions it affords. These representations are idiosyncratic and reflect only those actions that can be generated by the agent. Research has been done to investigate the autonomous acquisition of such affordance representations with intrinsic motivators. For instance, an example of multiple intrinsic reward functions have been proposed to learn the transition dynamics of a particular task [25]. Others have looked

into domain-independent intrinsic rewards, like novelty or certainty, for learning adaptive, non-stationary policies based on data gathered from experience [7, 29]. In particular, *model exploration programs* have been presented [7], but methods reported to date lack multimodal sensor integration and do not produce knowledge structures that are easily transferrable to other tasks.

A number of studies have presented methods to learn affordance representations through imitation [13, 14], building experience-grounded representations called *Object–Action Complexes* (OACs) [17]. Affordance models such as OACs provides a basis for *structural bootstrapping*, allowing existing knowledge to generalize to otherwise unexplored and novel tasks and domains [24]. Such generalizability may be used to support planners that use learned representations as *forward models* $f : a, s \mapsto s'$ which is synonymous to state transition models in MDPs. However these models do not necessarily encode the prerogatives of the embodied system nor can they be easily adapted to other robots [16]. Moreover, they require the guidance of a teacher and are relatively cumbersome, though methods have been proposed with intrinsic motivators but assume predefined structure prior to training [22].

This paper adopts a different form of affordance representation that is lighter weight, and thus, better serves planners that need to roll out a number of these forward models during planning. In fact, this representation encodes only essential Markovian components concerning information regarding states, actions, and transition dynamics, $s, a \mapsto s'$, and thus, can be reused generally for all tasks that can be formulated as MDPs. The main contribution of this paper is the presentation of an intrinsically motivated structure learning approach that builds complete action-related representations of objects using multimodal percepts. The result is called an *Aspect Transition Graph* (ATG) model. Previous planning architectures using hand built versions of these models have been successful, however, this paper contributes a structure learning approach to acquiring them autonomously.

We present the first autonomously learned ATG representation with continuously parametrized action edges in the literature. These representations can be used to serve as forward models in belief-space planning infrastructure on real robot systems [30]. A number of studies have integrated ATG affordance representations into the model base as a fundamental attribute in the model-referenced belief-space planning architecture. For instance, Sen showed that the

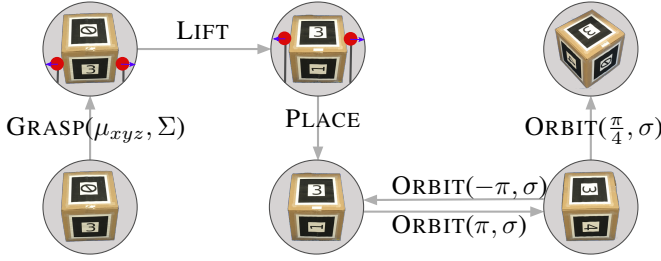


Fig. 1: An example of a *partially* constructed Aspect Transition Graph where affordances of a die-like object are encoded in *aspect nodes* connected by directed edges representing actions. The labels on the edges here correspond to possible *control programs* with parameters that result in successful transitions. The blue arrows at the hand for two of these nodes indicates tactile information.

*object identification task*¹ can scale up to 10,000 object models by pruning those with insufficient support [20]. Though models used in studies like those of Sen [18–21] do not inherently encode transition dynamics learned by the robot and therefore are not robust to unexpected outcomes without inherently encoding the system’s uncertainty into the representations. Work by Ku *et al.* has shown that the ATG structure is, however, capable of fine grain error detection, in which surprising outcomes that are not encoded in the object model trigger a *What’s up?* action [21]. Transition properties in our work are all learned autonomously, encoding unknown properties of the underlying system, and thus are robust to errors.

II. TECHNICAL APPROACH

This manuscript presents an algorithm for autonomous structure learning incorporating multiple sensor modalities and robot actions to produce lasting artifacts that can be used in the future (i.e. for reasoning and planning). The algorithm is presented in the context of a system that builds object representations for future use, but may span a number of other domains that require learning inherent structure in a task independent manner.

A. Affordance Representation

Our approach encodes affordances in a graphical structure called an *Aspect Transition Graph* [19]—which is defined as a directed multi-graph $G = (\mathcal{S}, \mathcal{A})$ where \mathcal{S} denotes a set of *aspect nodes* connected by action edges \mathcal{A} . A pictorial example of a partial ATG is illustrated in Figure 1 describing several plausible interaction outcomes with a particular object. Sensory information in multiple modalities (vision and touch) is integrated into the aspect nodes in the graph. Each parameterized action $a \in \mathcal{A}$ uses a learned search distribution for motor references that reliably transition between aspects. References are defined to be (multivariate) Gaussian distributions $\mathcal{N}(\mu, \Sigma)$ in Cartesian space describing the areas

¹*Object identification task*—a robot is given a large corpus of models it has interacted with in the past in memory and a real-world object and is asked to identify which model the object belongs to. Work of this nature falls into the *active vision* field which suggests that vision alone can not solve such tasks, but the embodied system must execute actions to condense its belief towards the correct object model in memory.

in object frame where the robot has successfully detected a target perceptual reference from this initial state in the past. An *aspect node* is a state representation defined as a geometric constellation of features derived from multiple sensor modalities. For example, an aspect node may be a geometric constellation of visual features present in a particular “field of view.” In theory, the number of features for any given object may be arbitrarily populous. As a result, the size of an ATG representation for the object, may be very large. However in principle, aspects encode *affordances* and are bounded by the number of actions $|\mathcal{A}|$ an embodied system may perform on the object that changes the relative sensor geometry. Aspect nodes are stored in the model as pointers to specific features that are indexed and arranged by type and value ordered chronologically by discovery time. Each feature is defined by three fields (id, type, value) in addition to a mean and covariance describing the likely Cartesian positions $\mu \in \mathbb{R}^3, \Sigma \in \mathbb{R}^{3 \times 3}$ in object frame².

B. Aspect Observation

Performing an observation of the scene creates a feature list. Observations consist of maximum likelihood Cartesian features derived from a Kalman filter that summarizes the history of observations to this point in terms of a mean observation and an associated spatial covariance. The current aspect s is obtained by observing the multimodal features in the scene and applying some mapping $\mathcal{F} : f_i, f_j, \dots, f_n \mapsto s$ which defines the aspect as a subset or encoding of relevant features in the feature list. This paper implements \mathcal{F} by simply returning the string representation of the geometric (order-specific) collection of all features over all modalities present, obeying some aspect geometry. Future work hopes to extend \mathcal{F} to incorporate a generalized Hough transform to vote for the position of the model coordinate frame [30].

C. Control Actions

Each edge in the ATG is a closed-loop controller $\phi|_{\tau}^{\sigma}$ that combines potential functions ($\phi \in \Phi$) with sensory ($\sigma \subseteq \Sigma$) and motor resources ($\tau \subseteq \mathcal{T}$) using the *Control Basis* framework [20, 31]. Such controllers achieve their objective by following gradients in the potential function $\phi(\sigma)$ with respect to changes in the value of the motor variables \mathbf{u}_{τ} , described by the error Jacobian $\mathbf{J} = \delta\phi(\sigma)/\delta\mathbf{u}_{\tau}$. References to low-level motor units are computed as $\Delta\mathbf{u}_{\tau} = \kappa\mathbf{J}^{\#}\Delta\phi$, where $\kappa > 0$ is a small gain, $\mathbf{J}^{\#}$ is the pseudoinverse of \mathbf{J} , and $\Delta\phi$ is the difference between the reference and actual potential. For the method proposed in this paper, it is assumed that the number of actions $|\mathcal{A}|$ and their parameter spaces are known *a priori*.

D. Action Selection

Affordances of a given object can be determined by exploring actions that cause a transition in the aspect space. The approach presented in this manuscript selects actions

²For instance, the bottom leftmost aspect node in Figure 1 could be defined as a 0-1 aspect, pointing to feature id: 0 of type: ‘ARtag’ and value: ‘3’ and feature id: 1 of type ‘ARtag’ and value: ‘0’.

to learn affordance representations—actions are selected for execution to achieve two kinds of reward:

- 1) **Discovering novelty** finding new aspect nodes $s \in \mathcal{S}$ and revealing new aspect transitions
- 2) **Refining parameters** $\rho \sim \mathcal{N}(\mu, \Sigma)$ for actions $a \in \mathcal{A}$ encoded in search distributions of known transitions $p(s'|s, a(\rho))$ between aspect nodes s and s'

To encourage coverage over the space of actions, a Latin Hypercube Sampled (LHS) space (over the domain of each action $a \in \mathcal{A}$) is introduced for each aspect node s . Action parameters ρ are randomly sampled during exploration—first by type, then by parameters defined in the LHS-space. *Parameter refinement* at some aspect node s is addressed by querying and evaluating all outgoing edges $\forall a \in \mathcal{A} : p(s'|s, a) > 0$ and sampling parameters ρ for the *highest-valued* action a^* . It is important to note that intrinsic reward alone is insufficient for producing a complete ATG representation since it selects actions that exploits learned structure, however, structure must first be discovered hereby stressing coverage.

E. Affordance Modeling and Intrinsic Reward

The *highest-valued* action a^* is a property of intrinsic motivation that drives the affordance modeling construction process. This is performed by storing and updating aspect node structure and transition information given learning experiences in the form of $\langle s, a, \rho, s' \rangle$, which is achieved by obtaining the aspect definitions s and s' (outlined in Section II-B) respectively before and after the selection and the execution of the action a with parameters ρ (from Section II-D). In short, the robot observes the current state, performs actions, and memorizes the new state produced. With each experience example, the parameter ρ is added to a nonparametric distribution along the action edge a in the ATG corresponding to the transition from s to s' . This nonparametric data structure stores all the robot's past experiences and describes all control parameters ρ that result in a particular perceptual outcome. Next, a new Gaussian distribution is generated using the set of all ρ existing in a and intrinsic reward $r(s, a(\rho), s')$ is computed.

Q-value iteration [32, 33] is implemented to establish a value function where at any time, the highest value corresponds to regions of interest in the parameter-space with high uncertainty. The intrinsic reward function uses the difference in the variance of experiences achieved from the same state under the same action as originally proposed in [7]. For use with the ATG representation, the intrinsic reward takes a slightly different form,

$$r(s, a(\rho), s') = \text{abs}(\|\Sigma_k\|_2 - \|(\Sigma_{k-1})\|_2)$$

where k indicates the value after the k th experiment, Σ_k refers to the sample variance in the Gaussian $\mathcal{N}(\mu, \Sigma)$ that describes the action parameter distribution that consists of all parameters ρ that accomplishes transition $s, a(\rho) \mapsto s'$. Then, Σ_{k-1} is the Gaussian distribution that does not include the most recent action $a(\rho_k)$. In the case that the edge a (corresponding to $s \rightarrow s'$) is novel or the aspect

s' is novel, a new transition edge or node is created—reward is then the differential variance between an arbitrarily wide Gaussian and an arbitrarily thin Gaussian centered at ρ . In practice, the reward is bounded by some defined maximum, for instance $r(s, a(\rho), s') = 1.0$. The key insight for using an intrinsic reward function of this form is that it encourages the *consumption* of reward through actions. In other words, it promotes the selection of actions that produce a high differential variance. As these distributions converge, intrinsic reward diminishes, hence encouraging other action parameters contributing to other transitions to be selected.

In addition to this update in the transition properties, all features that define the new aspect s' are either added or updated as appropriate. In summary the overall algorithm is described simply as,

Algorithm 1 Multimodal Structure Learning

- 1: $f \leftarrow \text{NIL}$
 - 2: **do**
 - 3: $a, \rho \leftarrow \text{Select params by LHS or } \arg \max_a f(s, a, s')$
 - 4: Do a, ρ and obtain experience $\langle s, a, \rho, s' \rangle$
 - 5: $r(s, a(\rho), s') \leftarrow \text{abs}(\|\Sigma_k\|_2 - \|(\Sigma_{k-1})\|_2)$
 - 6: Update value $f(s, a, s')$ with reward $r(s, a(\rho), s')$
 - 7: Update $\mathcal{N}_{s, a \rightarrow s'}$ with current action params $a(\rho)$
 - 8: **while** $f(s, a, s') > \varepsilon : \forall s, a, s'$
-

III. EXPERIMENT METHODOLOGY

A. Robot Platform

Experiments are done on a dynamic simulation of the uBot-6 platform, a 13 DOF, toddler-sized, dynamically balancing, mobile manipulator [34] equipped with an Asus Xtion Pro Live RGB-D camera (shown in Figure 2) and two ATI Mini45 Force/Torque sensors one in each hand (not shown in figure). Control actions are executed by the robot to establish new sensor geometries and reveal new aspects. Collectively, experimental results compile a total of over 250 hours of robot simulation.

B. Sensor Modalities (Features and Aspects)

The Asus RGB-D camera and ATI Mini45 Force/Torque sensors provides visual and tactile information to the robot. Visual features are extracted and ordered such that the feature list is populated with priority on feature Cartesian location (left to right, bottom to top of the image). Primitive tactile features consist of the contact force $\hat{\mathbf{f}} \in \mathbb{R}^3$, from which the sum of squared contact forces $\sum_{i=L,R} \mathbf{f}_i^T \mathbf{f}_i$ and sum of squared contact moments $\sum_{i=L,R} (\mathbf{r}_i \times \mathbf{f}_i)^T ((\mathbf{r}_i \times \mathbf{f}_i))$ are computed at the centroid of the pair of contacts measured, where L and R signify left and right, respectively. Bimanual grasp configurations where the squared force and moment residuals are minimized simultaneously are considered to be valid grasp hypotheses. This form of tactile information is added to the visual components of an aspect to obtain a multimodal *aspect node* which we propose as a representation of the *interaction state*.

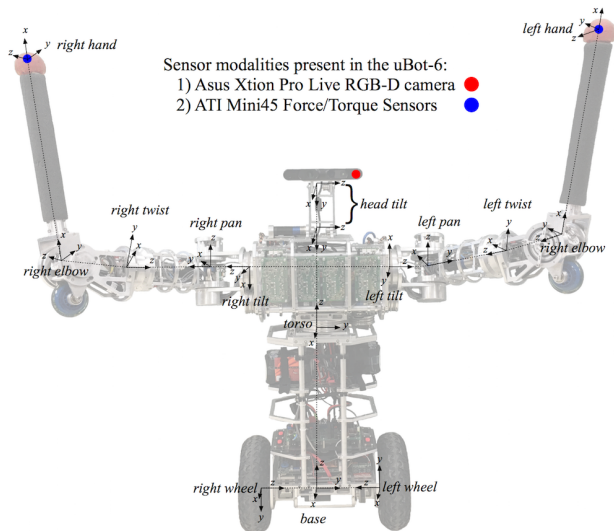


Fig. 2: The uBot-6 mobile manipulator has multiple degrees of freedom (DOF) supporting the ability to solve a number of tasks in many different ways. Control programs engage subsets of these DOF illustrated in the kinematic chain when executing actions in an attempt to search for intrinsic reward. Visual and tactile sensors allow for the robot to perform perceptual actions. (Best viewed in color)

C. Control Programs (Actions)

The set of actions \mathcal{A} in these experiments consists of two control programs: ORBIT and GRASP. These actions are responsible for changing the relative sensor geometries of the robot relative to the object and, as a result, cause probabilistic transitions to new aspect nodes.

ORBIT is a *locomotive control program* that changes the viewpoint geometry of the Asus Xtion Pro Live RGB-D camera. As shown in Figure 2, a number of motor resources (or degrees of freedom) may be used to achieve this. Our approach implements the translational and rotational axis of the base subject to the constraint that the final heading is toward the object. The action is parametrized by an angle θ about the world \hat{z} axis with some fixed orbit radius r defined a priori. In the experiment presented in this paper, $r = 1.0$ (m).

GRASP is a control program that changes the sensor geometry of the ATI Mini45 Force/Torque sensors in each hand to some new location in the scene. The grasp action engages the mobility resources, if necessary, to put the object within reach and then engages the arms to place the hands at Cartesian goals relative to the object where compressive forces are applied. This experiment is used to update the search distributions for mobility goals and hand placement based on the percept $\sum f_i = \sum m_i = 0$. A model is acquired that can be used to transform a partial observation of an aspect node, perhaps composed exclusively of visual features, into new aspect nodes that assert that the grasp force and moment objectives are also met (i.e. that an adequate grasp configuration exists). Each GRASP action is paired with a RELEASE counterpart in which the robot releases the held object and retreats to the previous pre-grasp base Cartesian position.

D. Target Objects

Since the approach presented in this paper makes no assumption regarding the underlying object and only concerns the aspects that are afforded, it can theoretically be applied to any object. However, in our validation experiments we use a simple object geometry as a proof of concept whose ATG can be evaluated. In all experiments, the uBot-6 is presented with an ARcube object in Gazebo with a random configuration. ARcubes are rigid 29 cm cubes with a single ARtag mounted on each of the six faces³. Visual observations of these features establish the location of the center of each tagged face.

The ARcubes in this experiment provides a form of validation since the number of aspects and transitions of ARcubes are enumerable by hand. Further, the transition parameters for ORBIT on this specific object tends to be at about an interval of $\pi/4$ —this intuition can be used to verify the correctness of models produced by our approach by establishing a ground truth representation.

E. Experiment Layout

The first experiment is a necessary validation step in which the proposed approach presented in this paper is compared against a base-line approach. Methods like [13, 14, 19] adopt either imitation or memorization paradigms for the construction of affordance models, with [19] being the only work specifically for ATGs. Work to learn ATGs have been mainly accomplished by learning through demonstrations, thus are not truly autonomous. Here, we present a base-line method in which the robot randomly explores control parameters, observes the scene, and memorizes its effects in terms of aspect transitions. Such a method is guaranteed to converge to a complete affordance model given sufficient time and serves as a valid contender for comparisons. Both the proposed and base-line methods are compared against a ground truth model with only the ORBIT action for validation.

The second experiment aims to inspect the result in which additional sensor modalities and actions are introduced. In the first experiment, the action space \mathcal{A} consisted solely of parametrized ORBIT actions and only the Asus camera existed in the set of sensor modalities. The uBot-6 has access to both ORBIT and GRASP actions and visual and tactile features from the Asus RGB-D camera and ATI force/torque sensors.

IV. RESULTS

The results presented in the first experiment contains over 150 hours of simulation, consisting of five trials for each approach. The affordance model corresponding to ground truth has eight visual aspect nodes and 64 interconnected transition edges in the ATG. Error (in radians) is computed by the absolute difference between the learned model and the ground truth for the means of the distribution along

³An open-source ARToolKit is available (<http://artoolkit.org>) for detecting and localizing the tags.

$ \mathcal{A} $	P-VALUE	PROPOSED	RAND+MEMORIZE
50	0.0034	1.4934 ± 0.0732	1.7551 ± 0.2097
100	0.0063	0.7598 ± 0.1477	1.0919 ± 0.2929
150	0.0068	0.4242 ± 0.1062	0.7666 ± 0.2876
200	0.0125	0.2701 ± 0.0630	0.5797 ± 0.3345
250	0.0071	0.2180 ± 0.0648	0.5474 ± 0.3435
300	0.0153	0.1574 ± 0.0409	0.4626 ± 0.3511
350	0.0217	0.1447 ± 0.0258	0.4388 ± 0.3511
400	0.0254	0.1251 ± 0.0187	0.4197 ± 0.3608
450	0.0251	0.1219 ± 0.0177	0.4071 ± 0.3511
500	0.0323	0.1112 ± 0.0060	0.3865 ± 0.3469

TABLE I: Model error comparison between the proposed structure learning approach and a base-line approach where transition probability from $s \xrightarrow{a} s'$ is approximated by randomly exploring action a from every initial state s and recording an estimated $p(s'|s, a)$ for all s with only the ORBIT control program in the action set. Evaluations are performed against an empirical model taken as ground truth and errors correspond to the average error (in radians) over all transitions in the model.

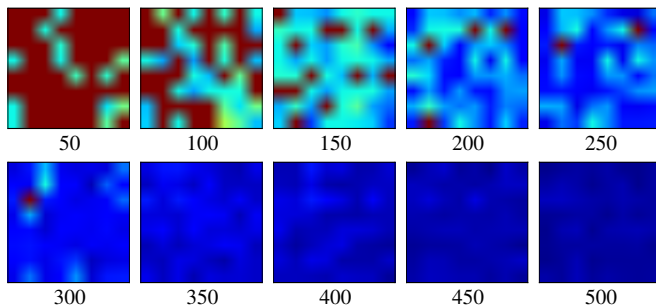


Fig. 3: Value functions of the same randomly selected run after a specific number of actions, indicating the current Q-values along state transition edges. Each row corresponds to a particular aspect node s and each column corresponds to the action $a(\rho)$ that results in a particular s' . Only ORBIT is considered in this image. The illustrated heatmap range is from (blue) 0.01 to (red) 1.0 (Best viewed in color).

all transition edges. The error related to each edge is then averaged for an overall model error. If an edge is not discovered by the learning method the error for that edge is set to the maximum, π . Table I lists the error for both the proposed and the random memorization approaches after a specific number of actions. In all cases, the proposed method achieves lower errors and in many of these cases, the difference is statistically significant ($p < 0.05$). It is also evident that the proposed approach is capable of acquiring more accurate affordance representations faster and more reliably (with significantly lower standard deviation).

As hinted in Section II-E, affordance structure learning does not converge to an optimal policy described by the resulting value function through value iteration. Instead, Figure 3 illustrates how the structure learning task *consumes* value and depletes the intrinsic rewards as the number of executed actions increase. These value functions illustrated the Q-values along transition edges corresponding to states s and actions $a(\rho)$. High values are attained initially when action parameters from aspect nodes have been inadequately unexplored. Immediate reward is consumed over time when actions are performed, condensing variance in transition distributions—such a phenomenon is illustrated in Figure 4 in which the log immediate reward depletes as learning progresses.

The introduction of additional sensor modalities and actions results in slightly slower convergence, yet continues

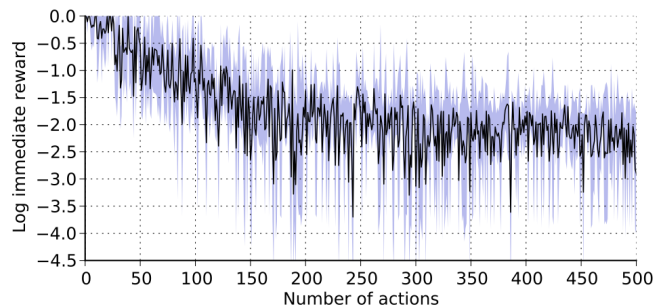


Fig. 4: Mean and one standard deviation over five trials for log immediate reward after the execution and update of each ORBIT action taken in the first experiment for the proposed approach.

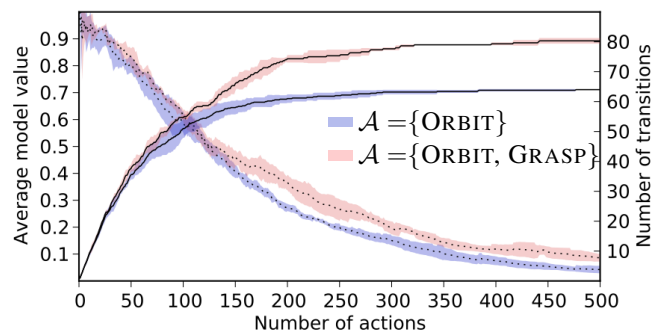


Fig. 5: Increasing sensor modalities and action space. Mean and one standard deviation over five trials illustrating the average Q value and the number of transitions discovered in the model using the proposed approach. The dotted lines correspond to average model value and solid lines describe the number of transitions in the affordance model (Best viewed in color).

to discover all the transitions in the learned ATG. The result of over 100 hours of simulation is illustrated in Figure 5. As the the number of transitions discovered in the model increases, the likelihood of novelty diminishes—this is captured in the decreasing values in the model. The affordance model with an extended sensor modalities and actions $\mathcal{A} = \{\text{ORBIT}, \text{GRASP}\}$ contains twelve aspect nodes and 80 aspect transitions. Like in the first experiment, structure learning with the extended action set requires 300–400 actions to produce a complete model. Other methods presented to learn affordances like OACs required a similar number of actions [22].

V. DISCUSSION, CONCLUSION, AND FUTURE WORK

This manuscript presents an intrinsically motivated structure learning approach to learn semi-permanent Markovian state representations of structures that are reusable in future (potentially partially-observable) tasks. The affordance representations learned here serves as a key component in belief-space object identification architectures [30]. These representations can be leveraged as forward models to predict how state distributions change in response to interaction. Despite success in the past using hand-crafted models of this type, the methods presented in this paper allows us to acquire them autonomously and encodes robot uncertainties and parameters that would otherwise be difficult to precisely hand define. Structure learning allows robots to build models themselves without supervision and promotes informed action selection, exploiting known structure and promoting a sense of discovery. Results demonstrate the acquisition

of models that are significantly better than approaches that solely select random actions to learn from. With the proposed learning method, the transitions encode uncertainties in the form of distributions derived from the properties of the embodied system and its interaction with the affordances of the world—as such, methods like this allow for a possibility for not only even finer-grained error detection, but also, support error *correction* in many cases, producing a more general and robust representation for planners and belief-space architectures. Future work looks into extending the action space and modalities further and investigating methods to take a learned affordance representation and decompose known aspect nodes to incorporate new sensory information while preserving learned transition dynamics. We believe that autonomously learning affordance representations as forward models with more complex actions and modalities allows for a richer set of future solvable tasks. Furthermore, despite the growing complexity of the affordance representation as more actions are introduced, we believe that enriched models reduce the complexity in model-referenced planning, thus reducing planning time and the number of rollouts necessary to solve future tasks.

ACKNOWLEDGEMENTS

The authors thank Michael W. Lanighan for his initial contributions. We also thank Mitchell Hebert and Samer Nashed for their feedback on this manuscript. This material is based upon work supported under Grant NASA-GCT-NNX12AR16A. Any opinions, findings, conclusions, or recommendations expressed in this material are solely those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration.

REFERENCES

- [1] N. Chentanez, A. G. Barto, and S. P. Singh, “Intrinsically motivated reinforcement learning,” in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. MIT Press, 2005.
- [2] L. Itti and P. F. Baldi, “Bayesian surprise attracts human attention,” in *Advances in Neural Information Processing Systems, Vol. 19*. Cambridge, MA: MIT Press, 2006, su.mod;bu;td;eye, pp. 547–554.
- [3] S. Singh, R. L. Lewis, A. G. Barto, and J. Sorg, “Intrinsically motivated reinforcement learning: An evolutionary perspective,” *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 2, pp. 70–82, June 2010.
- [4] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, “Intrinsic motivation systems for autonomous mental development,” *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 265–286, April 2007.
- [5] P. E. Utgoff and D. J. Straczuzi, “Many-layered learning,” *Neural Computation*, vol. 14, no. 10, pp. 2497–2529, 2002.
- [6] A. G. Barto, S. Singh, and N. Chentanez, “Intrinsically Motivated Learning of Hierarchical Collections of Skills,” in *International Conference on Developmental Learning*, 2004.
- [7] S. Hart, “An intrinsic reward for affordance exploration,” in *IEEE International Conference on Development and Learning*, June 2009.
- [8] N. J. Nilsson, “Shakey the robot,” DTIC Document, Tech. Rep., 1984.
- [9] P. K. Allen and R. Bajcsy, “Object recognition using vision and touch,” in *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, Los Angeles, CA, Aug 1985, pp. 1131–1137.
- [10] W. Grimson and T. Lozano-Pérez, “Localizing overlapping parts by searching the interpretation tree,” *PAMI*, vol. 9, no. 4, pp. 469–482, July 1987.
- [11] L. Natale, G. Metta, and G. Sandini, “Learning haptic representation of objects,” in *International Conference on Intelligent Manipulation and Grasping*, 2004.
- [12] A. Stoytchev, “Toward learning the binding affordances of objects: A behavior-grounded approach,” in *Proceedings of AAAI symposium on developmental robotics*, 2005, pp. 17–22.
- [13] M. Lopes, F. S. Melo, and L. Montesano, “Affordance-based imitation learning in robots,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2007, pp. 1015–1021.
- [14] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, “Learning object affordances: From sensory-motor coordination to imitation,” *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 15–26, Feb 2008.
- [15] L. Montesano and M. Lopes, “Learning grasping affordances from local visual descriptors,” in *International Conference on Development and Learning*, 2009, pp. 1–6.
- [16] R. Detry, D. Kraft, A. G. Buch, N. Kruger, and J. Piater, “Refining grasp affordance models by experience,” in *IEEE International Conference on Robotics and Automation*, May 2010, pp. 2287–2293.
- [17] N. Kruger, C. Geib, J. Piater, R. Petrick, M. Steedman, F. Worgotter, A. Ude, T. Asfour, D. Kraft, D. Omrcen, A. Agostini, and R. Dillmann, “Objectaction complexes: Grounded abstractions of sensorymotor processes,” *Robotics and Autonomous Systems*, vol. 59, no. 10, pp. 740 – 757, 2011.
- [18] L. Y. Ku, S. Sen, E. Learned-Miller, and R. Grupen, “Action-based models for belief-space planning,” in *Workshop on Information-Based Grasp and Manipulation Planning, at Robotics: Science and Systems*, Berkeley, California, July 2014.
- [19] —, “The aspect transition graph: An affordance-based model,” in *Second Workshop on Affordances: Visual Perception of Affordances and Functional Visual Primitives for Scene Analysis, at the European Conference on Computer Vision*, Zurich, Switzerland, Sept 2014.
- [20] S. Sen and R. Grupen, “Integrating task level planning with stochastic control,” University of Massachusetts Amherst, Tech. Rep. UM-CS-2014-005, 2014.
- [21] L. Y. Ku, D. Ruiken, E. G. Learned-Miller, and R. A. Grupen, “Error detection and surprise in stochastic robot actions,” in *15th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2015, Seoul, South Korea, November 3-5, 2015*, 2015, pp. 1096–1101.
- [22] E. Ugur and J. Piater, “Emergent structuring of interdependent affordance learning tasks,” in *Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics*, Oct 2014.
- [23] H. O. Song, M. Fritz, D. Goehring, and T. Darrell, “Learning to detect visual grasp affordance,” *IEEE Transactions on Automation Science and Engineering*, vol. PP, no. 99, pp. 1–12, 2015.
- [24] F. Worgotter, C. Geib, M. Tamosiunaite, E. E. Aksoy, J. Piater, H. Xiong, A. Ude, B. Nemec, D. Kraft, N. Kruger, M. Wchter, and T. Asfour, “Structural bootstrapping - a novel, generative mechanism for faster and more efficient acquisition of action-knowledge,” *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 2, pp. 140–154, June 2015.
- [25] T. Hester and P. Stone, “Intrinsically motivated model learning for developing curious robots,” *Artificial Intelligence*, pp. –, 2015.
- [26] E. Ugur and J. Piater, “Refining discovered symbols with multi-step interaction experience,” in *IEEE-RAS International Conference on Humanoid Robots*, Nov 2015.
- [27] P. Kaiser, M. Grotz, E. E. Aksoy, M. Do, N. Vahrenkamp, and T. Asfour, “Validation of whole-body loco-manipulation affordances for pushability and liftability,” in *IEEE/RAS International Conference on Humanoid Robots*, 2015, pp. 920–927.
- [28] J. J. Gibson, “The theory of affordances,” *Hilldale, USA*, 1977.
- [29] P. Sequeira, F. S. Melo, and A. Paiva, “Learning by appraising: an emotion-based approach to intrinsic reward design,” *Adaptive Behavior*, Sep. 2014.
- [30] R. A. Grupen, M. Hebert, M. W. Lanighan, T. Liu, D. Ruiken, T. Takahashi, and J. M. Wong, “Affordance-based active belief recognition using visual and manual actions,” in *IEEE/RSJ the International Conference on Intelligent Robots and Systems (IROS)*, Oct 2016.
- [31] M. Huber, W. S. MacDonald, and R. A. Grupen, “A control basis for multilegged walking,” in *IEEE International Conference on Robotics and Automation*, Apr 1996.
- [32] C. J. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, 1992.
- [33] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. Cambridge Univ Press, 1998, vol. 116.
- [34] D. Ruiken, M. W. Lanighan, and R. A. Grupen, “Postural modes and control for dexterous mobile manipulation: the umass ubot concept,” in *IEEE-RAS International Conference on Humanoid Robots*, Oct 2013.